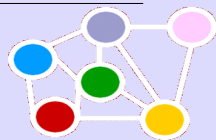
Digital Library Interoperability

technical and object modelling aspects

of

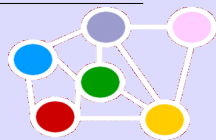
Europeana

Dr. Stefan Gradmann / EDLnet WP 2
stefan.gradmann@rrz.uni-hamburg.de
www.rrz.uni-hamburg.de/RRZ/S.Gradmann



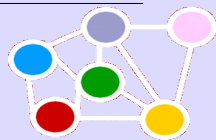
Overview

- Report on the work of EDLnet WP2 on "Technical and Semantic Interoperability" and relevant background
- Look back to Europeana context and vision
- Have a closer look at **Initial Semantic and Technical Interoperability Requirements** (D2.2)
  - Underlying **Digital Library Vision** and evolution of **Object Models**
  - **Technical Architecture**
  - With glances only at **Metadata** and **Multilingual Issues** as well as **Standards**
  - The way we intend to **interact with content providers**
- Short concluding message and pointer to the Final Session!



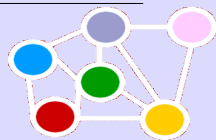
Europeana Context

- EC i2010 agenda with Digital Libraries as one of 3 'flagship initiatives': setting up the **European Digital Library** as a common multilingual access point to Europe's distributed digital cultural heritage including all types of cultural heritage institutions
  - **2008**: >= 2 million digital objects; multilingual; searchable and usable; work towards including archives.
  - **2010**: >= 6 million digital objects; including also museums and private initiatives.
  - *"I am **not** suggesting that the Commission creates a single library. I envisage a **network** of many digital libraries – in different institutions, across Europe." V. Reding (29 September 2005)*
- This last statement makes **Interoperability a key issue** for Europeana



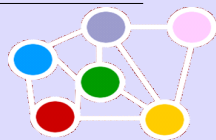
Why Interoperability?

- Europeana will be **federating** objects from **distributed sources**
- Europeana will be federating objects from **heterogeneous sources** with **different community background** – e. g. libraries vs. museums vs. archives ... but also scholars vs. policy makers vs. policy makers ...
- Europeana will be part of a bigger framework of **interacting global information networks** including e. g. 'Digital libraries', scientific repositories and commercial providers
- Europeana will have to be built with **minimal development efforts** and thus rely as much as possible on **standards** and **existing building blocks** as well as be **based on web standards**
- EC **WG on Digital Library Interoperability** active from January to June 2007 (More details on EC Working Group at http://bnd.bn.pt/seminario-conhecer-preservar/doc/Stefan%20Gradmann.pdf)
- And this is why interoperability has such a prominent place in the name of the "technical" WP of EDLnet: **Interoperability is the heart of the technical vision of Europeana**!

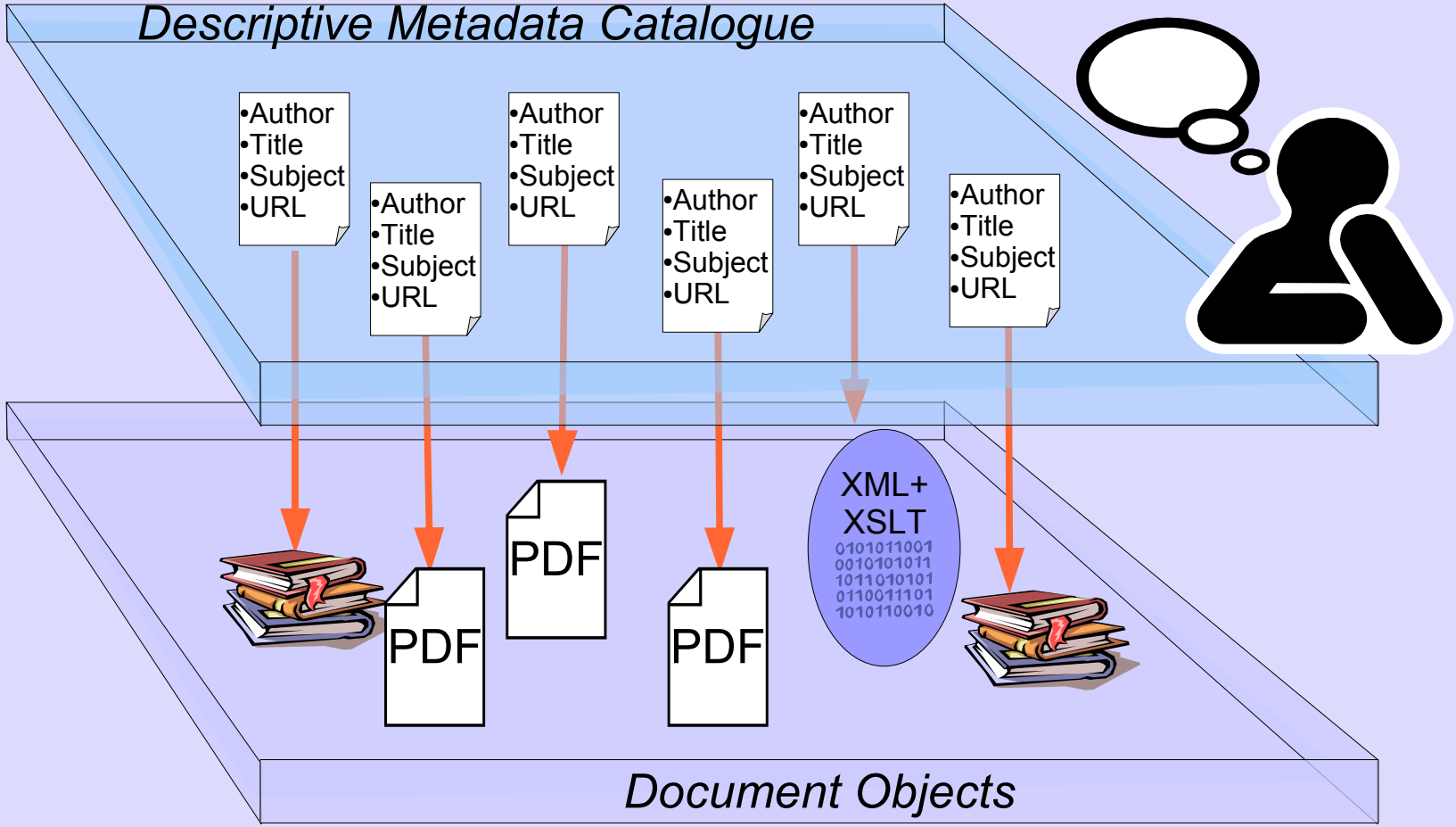


From EC Working Group to EDLnet WP2

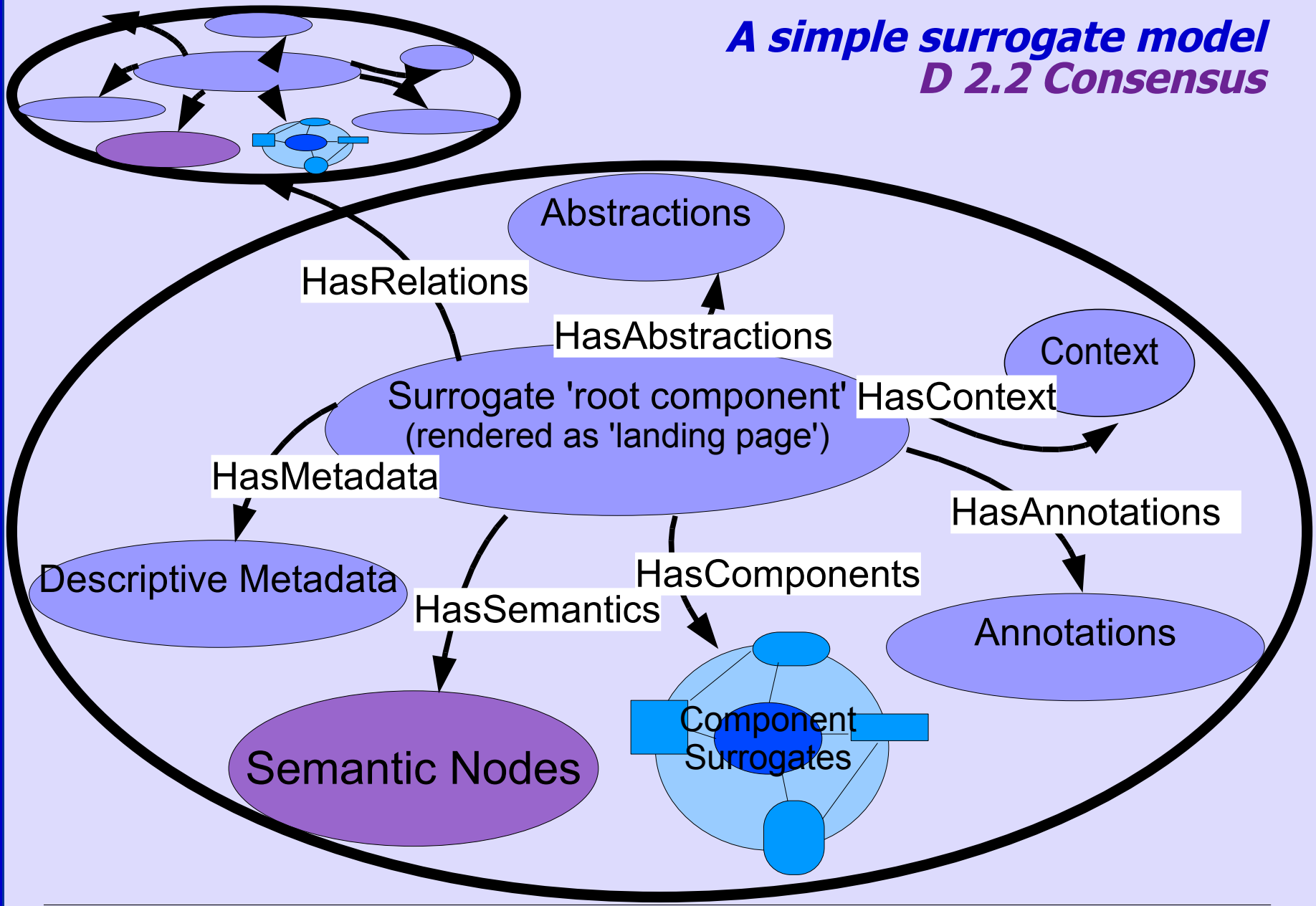
- Both Short Term Agenda Issues and conceptual framework were input for the **WP2 Working Groups** of EDLnet:
  - WG 2.1 Standards & Interoperability (Makx Dekkers)
  - WG 2.2 Semantic and Linguistic Interoperability (Stefan Gradmann)
  - WG 2.3 Technical Interoperability (Carlo Meghini)
- Work from 09/2007 to 12/2007 was concentrated on **Initial Semantic and Technical Interoperability Requirements (D2.2)** which is both short term oriented (with the Europeana prototype as target) and 'visionary' (feeding into the Europeana 'maquette')
  - Big picture (conceptual switch as mentioned by Herbert, starting from the 'object' concept: resource driven)
  - Component architecture
  - Metadata and Multilingualism
  - Interaction with content providers



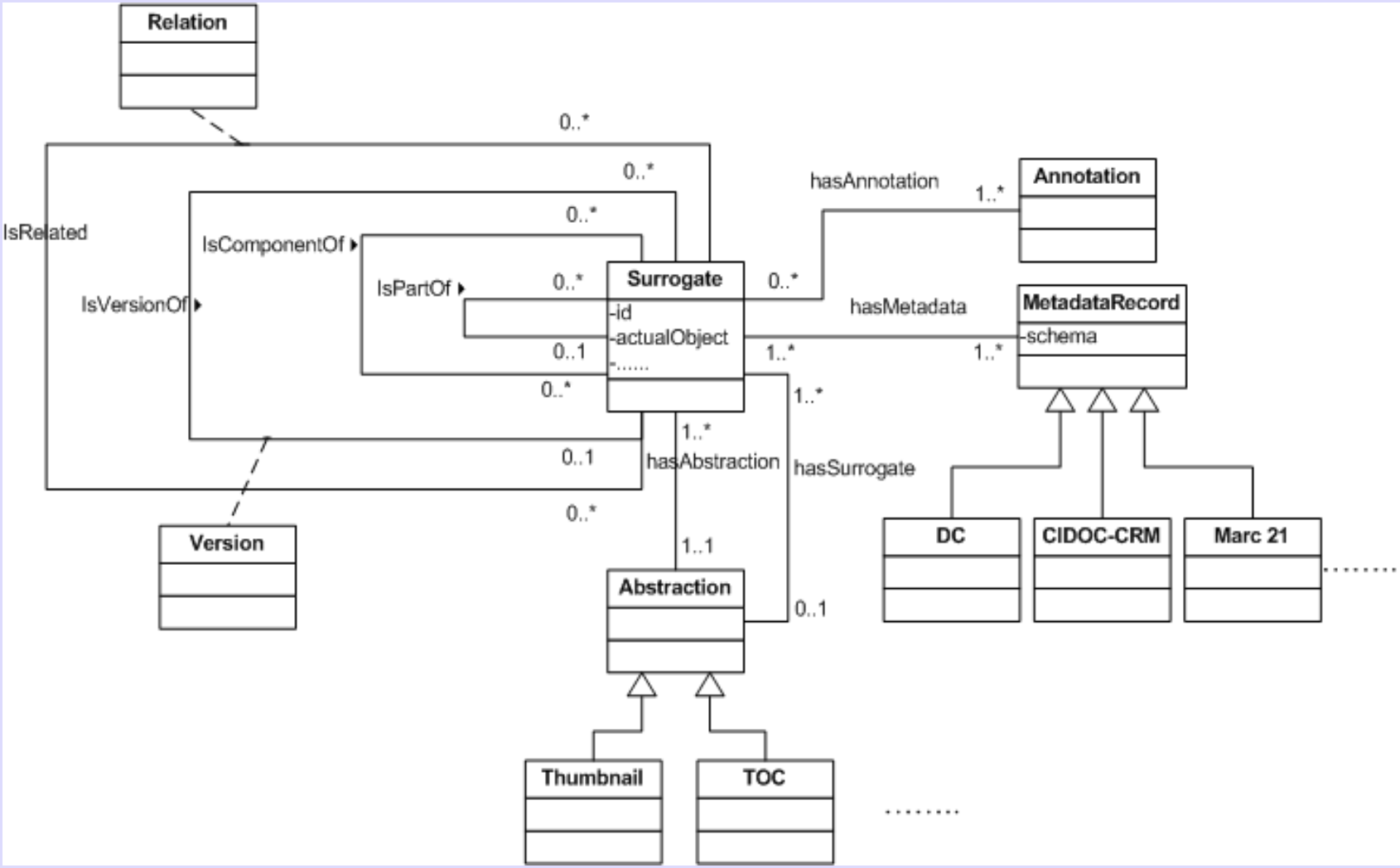
Metadata and Objects
In catalogue based (digital) libraries



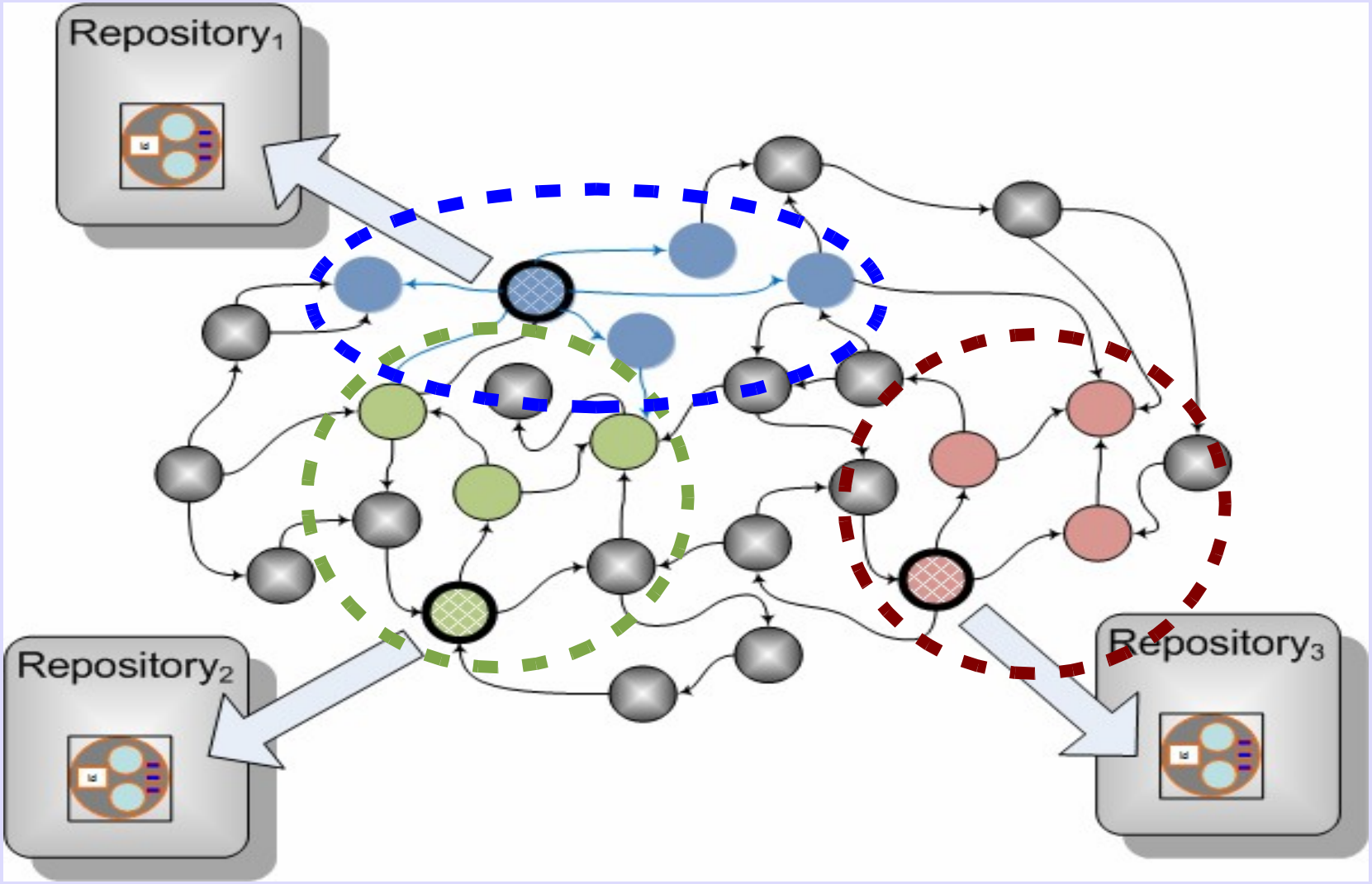
A simple surrogate model
D 2.2 Consensus



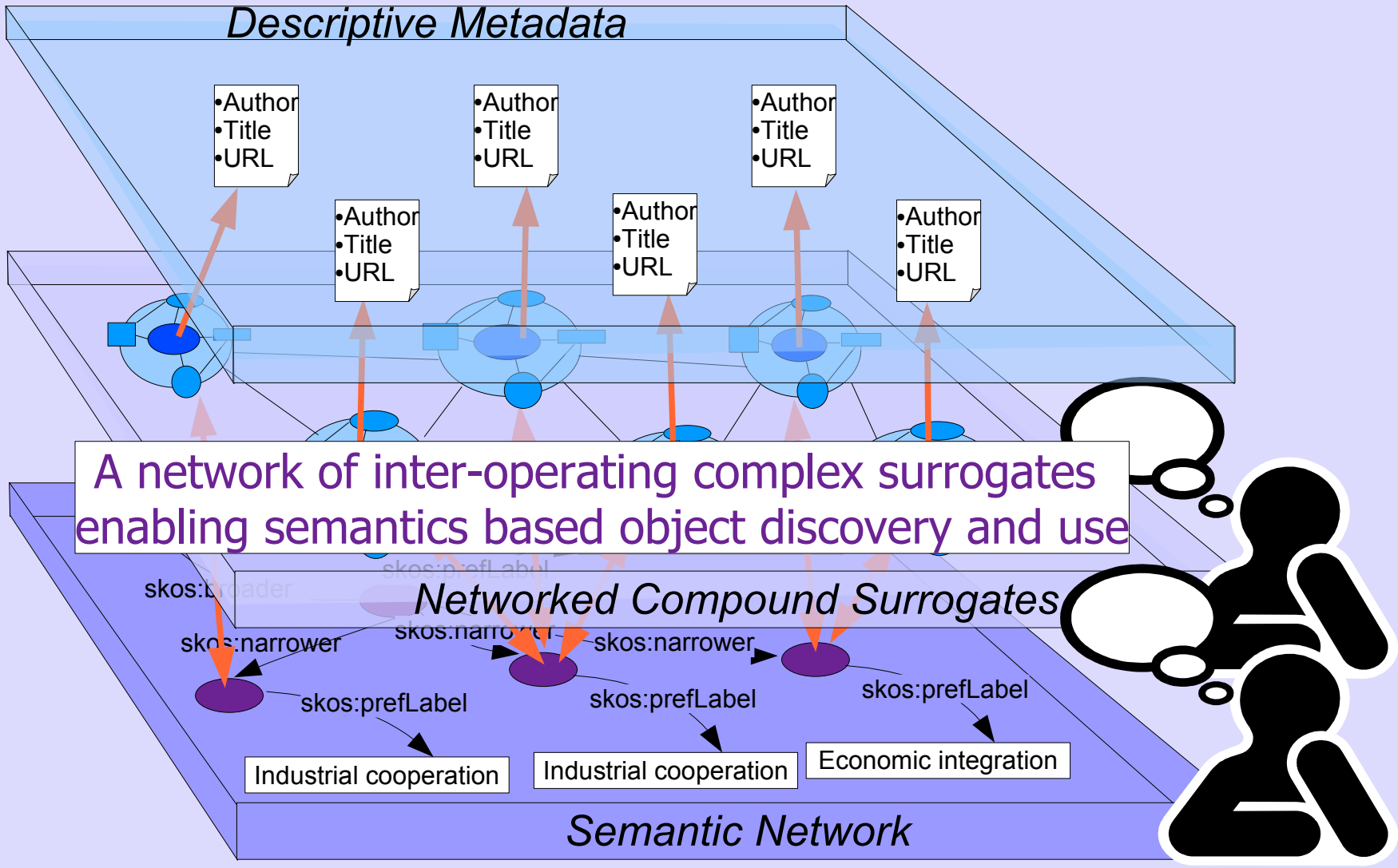
A simple surrogate model
First Data Model Formalisation



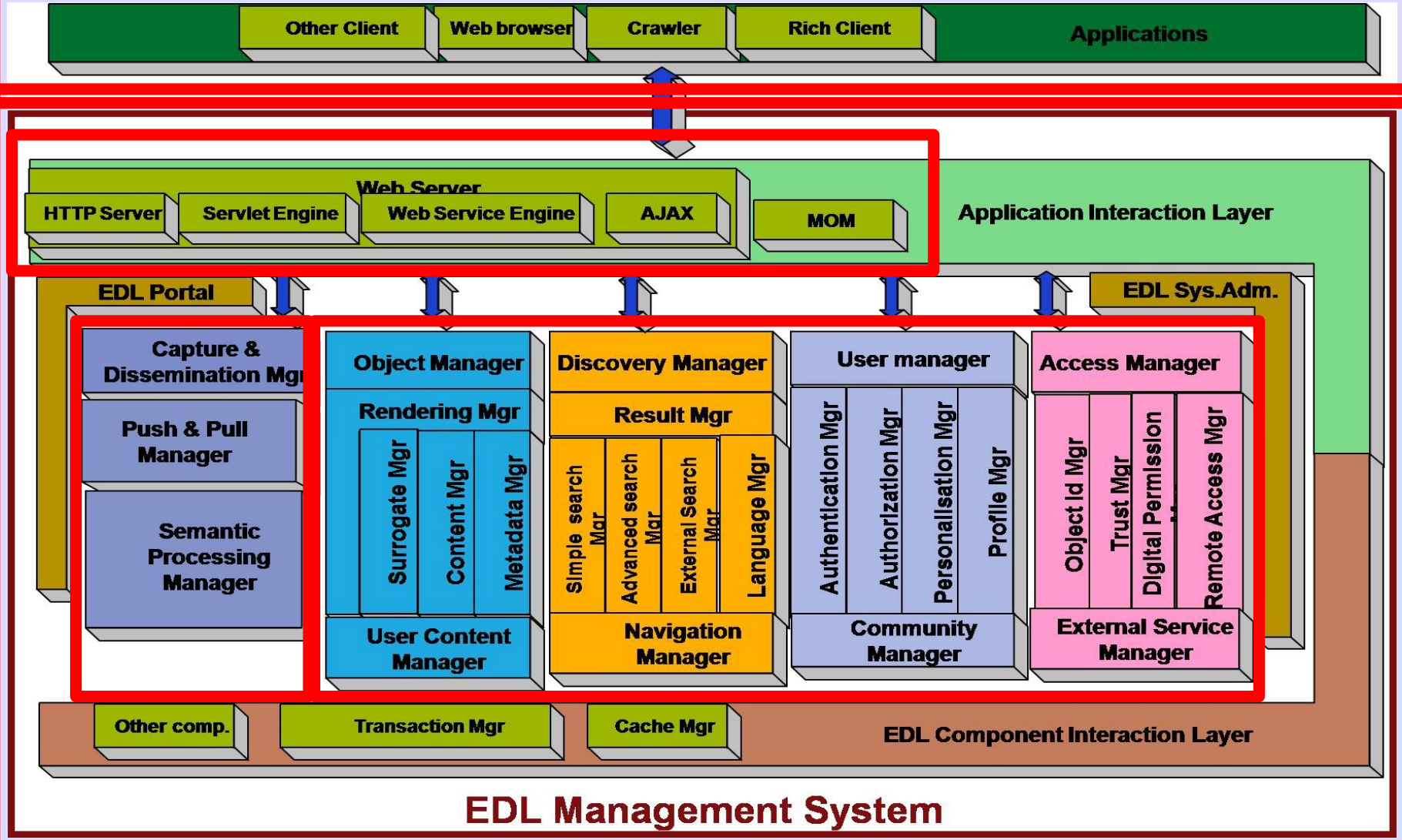
A complementary and more granular model
Object Reuse and Exchange (ORE)



Document Objects, Metadata and Semantic Networks

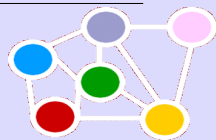


Europeana Component Architecture



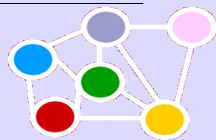
Some Words on Metadata and Multilingual Issues

- **Metadata**:
  - OAI-PMH (Header, Metadata, About)
  - DC unqualified (occurence, type, encoding and vocabulary specified)
  - Other formats are optional and a limited number only should be used with the *metadataPrefix* => used for full text searching
  - Use of semantic interoperability techniques for semantic mappings and the cross-searching of descriptive metadata **instead of** a higher level interoperability application profile.
- **Multilingual Issues**: 4 levels for implementing MLIA
  - Application User Interface
  - browsing via a common multilingual ontology mapping onto versions for each language
  - Search on a monolingual baseline (for all languages supported (i) and simple cross-language search using query translation (ii))
  - Full multilingual search & presentation – certainly not in prototype!



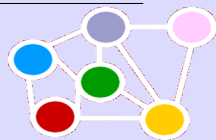
Interaction with Content Providers and End Users

- **Content providers** need to provide
  - **identifiers**, **metadata files**, **vocabularies** in SKOS form, **links to semantic nodes**, **licensing and rights information** and **temporary access** to the original digital objects (skipping details here)
  - For all objects that are referenced by Europeana surrogates, **a URL that points to the binary original object**
  - **Standard identifiers for digital objects**, and the content providers should be responsible for persistent resolution.
- The provided elements will be **aggregated into Europeana surrogates** which in turn are the basis for Europeana functionality
- Access for end users to the original binary objects is via a link to the content provider site as part of the surrogate: **Europeana will not store original objects**!



Some Words on Standards

- **OAI-PMH** for harvesting
- **SRU**, **SOAP**, **JSR** and **OpenSearch** for external search and access
- **XML** for syntax
- **SKOS**, **OWL/RDF(S)** for semantics
- **SPARQL** for advanced search (but still discussing alternatives such as SQL, SeQL and Lucene)
- **CIDOC/CRM** and **DCMI Abstract Model** Metadata modelling (**FRBRoo** to be discussed)
- **SAML**, **LDAP** and **OpenID** for security and authentication
- Lots of others: WSDL, WAI, UNICODE UTF-8, XSLT, NACO, UN-API "See-also", HTTP/HTTPS ...



Where are we now?

- We have initial requirements / ideas of how to inter-operate
  - On **object** level
  - On **metadata** level
  - On **component** level
  - In terms of **multilinguality**
  - With **content providers**
- And we have the 'maquette' giving a first idea of how all this may be perceived from an **end user perspective**
- As for the **road ahead**: see you again ~14:15.
- For questions: please contact me directly under stefan.gradmann@rrz.uni-hamburg.de or as well Julie Verleyen under Julie.Verleyen@KB.nl